DOCUMENT RESUME

ED 421 506 TM 028 465

AUTHOR Mills, Rollin W.
TITLE Development of Program and Individual Student Evaluation Models for Foreign Language in the Elementary School.
PUB DATE 1998-04-00
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Difficulty Level; Elementary Education; French; Japanese; *Language Tests; *Listening Comprehension Tests; Models; Program Evaluation; Reading Tests; Second Language Instruction; Second Language Learning; *Second Language Programs; Spanish; Student Evaluation; Tables (Data); Teacher Made Tests; *Test Construction
IDENTIFIERS South Carolina

ABSTRACT

This study was conducted to evaluate the tests developed by elementary foreign language teachers of French, Japanese, and Spanish in a school district in South Carolina. The tests were designed to determine the level of end-of-year student learning and to provide a basis for evaluating the curriculum of each of the three languages. The French and Spanish tests contained tests of listening and comprehension, vocabulary, and reading, and the Japanese test contained tests of listening, complex listening skills, and vocabulary. The tests were analyzed in terms of item difficulty, high-low discrimination indices, and distributions patterns. The subtests were also analyzed, highlighting the tendency of teacher-made tests toward the measurement of minimal skills. The study provides descriptive statistics for all parts of the tests and the total test results. Analysis indicates that, in general, all three tests had too low a level of difficulty, with few questions to challenge the more able students. These results are a contribution toward the improved design of foreign language tests for elementary school students, for whom foreign language study is still relatively rare. (Contains 9 tables, 30 charts, and 17 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 421 506

Development of Program and Individual Student Evaluation Models for Foreign Language In the Elementary School

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Rollin Mills
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Dr. Rollin W. Mills

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in San Diego, CA
April 1998

TM028465

Development of Program and Individual Student Evaluation Models for Foreign Language In the Elementary School

Presented at

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

Annual Conference

April 13-17, 1998

Session 27.40

Wednesday, April 15, 1998, 2:15-3:45 P.M.

Hyatt Hotel

Room Windsor A, 3rd Floor
San Diego, California

by

Dr. Rollin W. Mills
Coordinator of Educational Assessments
Lexington School District One
Lexington, SC
rmills@lex1.k12.state.sc.us

Development of Program and Individual Student Evaluation Models for Foreign Language in the Elementary School

(Abstract)

This study was completed to evaluate the tests developed by elementary foreign language teachers of French, Japanese and Spanish. The instruments are designed to determine the level of end-of-year student learning and to provide a basis for evaluating the curriculum of each of the three languages. The tests are divided into three parts: (1) Listening and Comprehension; (2) Reading, except Japanese; and (3) Vocabulary. The second part of the Japanese test consists of complex listening skills.

The study provides an analysis of these instruments in terms of item difficulty, high-low discrimination indices and distribution patterns. Attention is also given to the three specific sub-tests of each diagnostic instrument with particular consideration to differences in distribution patterns among the tests. The study highlights the tendency of teacher made instruments toward the measurement of minimal skills.

The study provides descriptive statistics for all parts of the tests and for the total test results. In addition, discrimination and difficulty indices were completed and scatter plots created in relationship to item numbers. Analysis of results indicated that in general all three tests had too low a level of difficulty. There were few questions that challenged the more able students. Difficulty levels of items appear to be independent of item location in the tests.

This investigation provides the information necessary to assist in determining needed changes for improving the discrimination ability of the tests making them more reliable and valid instruments for use in foreign language curriculum and program evaluation. Since elementary foreign language programs are fairly new and require much time and effort to implement, it is important that educators develop effective evaluation tools to assist in making curricular program improvements. This study is a first contribution toward that end.

Table of Contents

Chapter 1	Introduction		3
Chapter 2	Review of Literature		5
Chapter 3	Development of the Evaluation Model		10
	3.1	Purpose and Rationale	10
	3.2	Procedures Utilized in Instrument Development	10
	3.3	Description of the Instruments	11
Chapter 4	Analysis of Results		13
	4.1	French	13
	4.2	Japanese	19
	4.3	Spanish	25
Chapter 5	Recommendations for Revisions and Further Research		31
	5.1	Needed Revisions to Testing Program	31
	5.2	Recommendation for Further Study	32
References			34

Chapter 1

Introduction

In 1990, the district began its elementary foreign language program in one school with approximately 150 students and a part-time French instructor. During the 1997-98 school year, with funding from the local school board and district administration and a federal grant entitled "Bringing the World to the Midlands", the program serves over 7,500 students in grades 1-6. The district administration is seeking additional funding to expand the sequential study of French, Japanese, and Spanish through grade 12 with students having the option of taking Advanced Placement offerings at the high school level.

The district has the only foreign language program of its kind in South Carolina. The district provides intensive staff development in FLES (Foreign Language in the Elementary School) methods and assists other school districts within South Carolina, including three of the more prominent midlands school districts, who are currently striving to establish elementary foreign language programs.

Staff members work as mentors for other districts to facilitate the introduction of FLES programs. A special FLES intensive workshop is held for new teachers at the University of South Carolina with district teachers teaming with the university's faculty members to provide this valuable training. South Carolina does not have a teacher certification program for elementary foreign language. There is an urgent need to provide as much training as possible for new teachers. This joint collaborative training is an effort toward that end.

The district foreign language teachers constructed an end-of-year test to be given to students enrolled in all three language courses. The purposes of the test were two-fold. The primary purpose was to determine which students taking the test had acquired enough basic vocabulary and understanding of the language to move into an accelerated foreign language track during the next school year. The second purpose of the test was to assist the district administration in determining if the curriculum common to all three languages had been delivered in a uniform manner by all foreign language teachers.

Each foreign language test consisted of three parts: Listening Comprehension, Reading and Vocabulary; except Japanese Part 2 which involved more complex listening skills than the skills tested by Part 1. The scores on all three parts were added to obtain a total score.

There are several purposes for this study. First, the primary purpose is to provide an analysis of the test results to use in evaluating the effectiveness of the elementary foreign language programs in French, Japanese and Spanish. Another purpose is to utilize the evaluation of the quality of test items to determine necessary revisions by the foreign language teacher committee in the test questions for the ensuing school year. A third purpose is to lay the groundwork for future studies by providing a model for other districts to follow in analyzing end-of-year test results in their respective school districts.

Chapter 2

Review of Literature

Currently much attention is being given by school districts to *Standards for Foreign Language Learning: Preparing for the 21st Century*, a booklet published by the American Council on the Teaching of Foreign Languages, American Association of Teachers of French, American Association of Teachers of German, and American Association of Teachers of Spanish and Portuguese. This organization has developed basic learning standards for foreign language at the elementary and secondary levels. The standards reflect five education goal areas: Communication Skills, Understanding Cultures Associated with Languages, Interconnectedness of Language and Other Bodies of Knowledge, Comparisons that Offer Insight Into the Nature of Language and Culture, and Participation in Multi-Lingual Communities. Thirty-four sample "learning scenarios" in which classroom activities reflect the standards are described in the document. The booklet also contains a list of frequently asked questions concerning the teaching of foreign language (American Council on the Teaching of Foreign Languages, 1986).

Although most districts are attempting to follow these national standards, the methods of implementing foreign language curriculum vary greatly from school district to school district throughout the United States. Also, school districts are at various stages of planning and implementation of foreign language programs. The extent and effectiveness of these programs are affected by

issues such as budget, commitment of the community and parents, geographical location, and political philosophy.

There are several arguments based on brain research for teaching foreign language in the elementary school. Recent studies of PET scans (positron emission tomography) show that by age four children's brains are twice as active as adult brains. The higher level of activity is because a child's brain maintains trillions of connections between neurons, double what will eventually be kept. Patti Mantrel, author of one of these studies, adds, "Synapses or avenues in the brain are opened up by foreign language instruction when it is introduced at an early age. If languages are not introduced at an early age, these synapses are not accessed, and language learning is much more difficult to acquire in later years" (Foreign Language and Youth, 1996).

In a recent article in *Technology Review,* Michael Phelps, a UCLA biophysicist and co-inventor of the PET scan, said, "The thing that determines which connections are saved is education in the broadest sense of the term. If we teach our children early enough, it will affect the organization, or 'wiring' of their brains." Phelps also noted that children can learn to "think" in a foreign language because their brains have the extra connections. Teaching a foreign language to young children provides the benefit that their brains will retain connections to the cerebral cortex that will enable them to better use and retain the foreign language (Foreign Language and Youth, 1996).

The editor of *The Times Record* of Brunswick, Maine states in a September 9, 1997 editorial that young children learning a foreign language are

willing to take risks and take part in such instructional methods as games, nursery rhymes and songs—learning by doing. They can learn more quickly; their pronunciation is better; and they remember what they learn. Specialists say that speech areas of the brain are firmly established after age ten to twelve, and the optimum time for learning a foreign language expires (Learning at an Early Age, 1997).

While a number of states are beginning to require the teaching of foreign language in the elementary school, there appears to be little consistency in the instructional programs offered. For example, in Maine, Falmouth School District has begun offering twenty minutes of French instruction per day to first graders and plans are to expand the program by one grade per year until languages are taught in all grades 1-12. Just 7.5 percent of Maine's public schools offer any foreign language to elementary school students (Learning at an Early Age, 1997).

In North Carolina, a number of school districts are now offering foreign language programs to elementary school students. Some kindergarten classes are also participating in the foreign language program. Most of these programs have fairly well-developed goals such as building cross-cultural understanding, developing communication skills through listening, speaking, reading and writing, and expanding the students' knowledge of math, science, language arts, social studies, and cultural arts. Catawba County has had a FLES program since 1988 and serves as a resource to other school districts throughout the state (North Carolina Department of Public Instruction, 1995).

A number of school districts and schools throughout the country are offering foreign language exploratory programs (FLEX) in the elementary schools. Most of these programs simply give students a foundation for foreign language study and assist them in deciding if they would like to take a foreign language later in their school career. Most of these courses are taught by itinerant or regular classroom teachers and are not a part of a foreign language program with in-depth language instruction.

Since the nature of the instructional programs and content are vastly different from district to district, assessment procedures for determining the effectiveness of the established programs also differ. Most district-wide assessment measures were constructed to assist in evaluating a specific program. Current assessment instruments tend to be very narrow in scope, often measuring one area such as the effects of a particular technique or procedure used by the program. Several examples illustrate this tendency.

Vivas in "Language Learning" reports on an experimental investigation of the effects of a systematic story-reading aloud program on student learning. Study results indicated that students increased their language comprehension and expression when listening to stories read aloud (Vivas, 1996).

Julia Henley in "Using Video as an Advanced Organizer to a Written Passage in the FLES Classroom" compares the effects of two visual advanced organizers on comprehension and retention of a written passage in a FLES classroom. The uses of video and pictures plus teacher narrative were

compared. Video was found to be the more effective organizer of the two (Hanley, 1995).

Richard Donato reports on a three-year study involving the comparison of two strands of research. The first strand deals with community and school ambiance and attempts to capture the attitudes and perceptions of parents, teachers and students. The second strand investigates the oral language achievement of children, focusing on oral proficiency, vocabulary development, and social uses of languages. Results indicate that over a three-year period, all children can make considerable progress in foreign language proficiency and develop positive attitudes toward learning (Donato, 1996).

Another study by May Hancock dealt with student perceptions and attitudes concerning the elementary foreign language program. Student comments were solicited on the strengths and weaknesses of the program. Study results were used in an attempt to improve the program's design and content (Hancock, 1995).

There are a few tests that have been developed on pronunciation of key words in the various foreign languages but there are no published tests and research on end-of-grade measures that are used in evaluating the total foreign language curriculum of a school district. There is a tremendous need for further research and data on the effectiveness of the various models being used by school districts throughout the country to deliver instruction in the elementary school.

Chapter 3

Development of the Evaluation Model

Purpose and Rationale

The implementation of a new elementary foreign language program necessitates the evaluation of both student progress and program effectiveness. The absence of any meaningful evaluation instrument mandates the development of this critical component. This study encompasses the development, analysis, and revision of student assessment instruments for each of the three languages: French, Japanese, and Spanish. The developed instruments possess a great deal of content validity because they were constructed by the teachers who provide the foreign language instruction. However, reliability and construct validity are less clear. It is the purpose of this study to analyze these instruments in sufficient detail so that the information provided will guide the revision and improvement of the instruments and enhance their role in the formative evaluation of the foreign language program.

Procedures Utilized in Instrument Development

Initially, all foreign language teachers had input into the writing of assessment items. They met as a group to suggest and/or write the items. Then a committee of teachers representing all three languages was selected to complete the item development process.

The committee studied the construction and content of similar tests developed by other states, school districts and professional organizations. An

attempt was made to determine the pieces of foreign language in those districts' programs that committee members believed necessary for all children to master. The committee met with the Foreign Language Coordinator from the South Carolina Department of Education to construct an initial draft of the test utilizing the ideas and suggestions discussed at previous meetings.

The initial drafts of the instruments designed during the meeting were sent to all foreign language teachers for feedback. The committee then met several times to make revisions to the initial drafts. The instruments were designed so that as little English as possible was used in the directions. Directions were given in the foreign language of each instrument in Part 1, and pictures were used for answers in Part 3. The pilot tests were administered during March, 1997.

Description of the Instruments

Each instrument is divided into three parts: (1) Listening and Comprehension; (2) Reading (except for Japanese which utilized more complex listening skills); (3) Vocabulary. Part 1 contains 30 questions in French and Spanish and 25 questions in Japanese. Part 2 contains 10 questions for all 3 languages, and Part 3 contains 15 questions for French and Spanish and 10 questions for Japanese. The result is a total of 55 items on the French and Spanish instruments and a total of 45 items on the Japanese instrument. All questions are multiple choice with responses recorded on a bubble sheet to

facilitate machine scoring. Students have 25 minutes to complete Part 1 and an additional 25 minutes to complete Parts 2 and 3 together.

In Part 1 of the instruments, students listen to the teacher read a statement and then identify a picture which matches the verbal cues provided. In Part 2 of the French and Spanish Instruments, student read a post card and answer questions concerning statements made on the card. Part 2 of the Japanese test contains additional, more difficult listening questions based on verbal prompts. Part 3 requires students to match a picture with commonly used words or phrases. These questions cover a wide range of vocabulary with which students should be familiar such as time of the day, feelings, numbers, and colors.

Chapter 4

Analysis of Results

French

Normalcy tests indicated deviation from a normal distribution of scores at the .05 probability level. However, the deviations tended to be limited to the upper end of the distribution where clearly the number of high scores exceeded normal distribution expectations.

The descriptive statistics for French test results by subtest and total score are shown in Table 1. A total of 350 students took the French test. The total scores ranged from 14 to 55 with several students scoring a zero on Part 2 and Part 3 of the test. The mean Total Score was 37.4.

Table 1: Summary Descriptive Statistics for French by Subtest and Total

Variable	N	Mean	Median	Tr Mean	Min	Max	St Dev	SE Mean
Part 1	350	20.2	20	20.3	3	30	5.414	0.289
Part 2	350	5.5	5	5.5	0	10	2.816	0.151
Part 3	350	11.7	12	12.0	0	15	2.987	0.160
TOTAL	350	37.4	37	37.5	14	55	9.771	0.522

Correlations were obtained among Part 1, Part 2, Part 3 and Total Test scores. The parts are less related to each other than to the total test with a range of correlations between parts and the Total Test ranging from a low of 0.80 to a high of 0.93. Correlations among the parts ranged from 0.587 to 0.663. These correlations were as expected since the parts of the test assessed different language skills. See Table 2.

Table 2: Correlations Among Parts and Total for French

	Part 1	Part 2	Part 3
Part 2	0.663		
Part 3	0.595	0.587	
TOTAL	0.927	0.835	0.804

The Split-Half Correlation for the Total Test was 0.843 with a resulting Total Test Reliability of 0.915.

Difficulty and Discrimination Indices were computed for each item. These indices are found in Table 3. This information is also depicted graphically in Chart 1 and Chart 2. Defined according to Classical Testing Theory, the Difficulty Index can actually be described as the "easy" index. The higher the index number, the easier the item. The Item Difficulty Index for French ranges from 0.40 to 0.98. There are 30 test items, out of the 55, with a Difficulty Index of 0.65 to 0.98. Overall, there appear to be too many easy test items.

The Discrimination Index (Percentage of high-scoring group getting item correct – Percentage of low-scoring group getting item correct) ranges from –0.02 to 0.63. Seventeen items have a Discrimination Index below 0.3 and six items are below 0.2.

Table 3: French in the Elementary School
Item Difficulty and Discrimination

Item	Diff Index	Discrim Index	Item	Diff Index	Discrim Index	Item	Diff Index	Discrim Index
1	0.54	-0.02	20	0.57	0.40	8	0.49	0.36
2	0.67	0.40	21	0.58	0.48	9	0.40	0.45
3	0.91	0.18	22	0.67	0.27	10	0.51	0.37
4	0.65	0.33	23	0.82	0.23	**Part 3**		
5	0.60	0.29	24	0.51	0.23	1	0.89	0.19
6	0.50	0.38	25	0.65	0.30	2	0.76	0.35
7	0.53	0.37	26	0.67	0.27	3	0.78	0.36
8	0.82	0.29	27	0.84	0.20	4	0.53	0.40
9	0.64	0.24	28	0.76	0.33	5	0.69	0.35
10	0.75	0.22	29	0.98	0.00	6	0.61	0.40
11	0.70	0.19	30	0.67	0.29	7	0.55	0.35
12	0.66	0.37	**Part 2**			8	0.92	0.13
13	0.82	0.33	1	0.71	0.29	9	0.94	0.10
14	0.74	0.23	2	0.42	0.48	10	0.75	0.32
15	0.48	0.27	3	0.53	0.58	11	0.85	0.26
16	0.49	0.32	4	0.57	0.63	12	0.78	0.33
17	0.64	0.44	5	0.61	0.44	13	0.91	0.15
18	0.51	0.30	6	0.69	0.43	14	0.89	0.15
19	0.86	0.24	7	0.55	0.29	15	0.88	0.16

Score distributions for Part 1 are shown in Charts 3 and 4. Charts 5 and 6 show distributions for Part 2 and Part 3. Chart 7 shows the Total Score Distribution by range of scores.

The scores for Part 1 ranged from 3 to 30 with 28 people scoring 21. The distribution pattern for Part 1 is fairly normal. When scores are grouped in 5-point ranges, 106 students scored between 16 and 20 and 72 students scored in the 26-30 range.

There are only 10 questions in Part 2. The distribution is bimodal. While 7 students scored a 0, the majority of scores clustered in the ranges of 3-6 and 8-10.

The Part 3 distribution is very negatively skewed with 67 students making the maximum score of 15. Of 350 students tested, only 12 scored 5 or below.

The Total test scores grouped by ranges of 5 tend to be fairly normally distributed with greater numbers of students scoring above the mean than below the mean. The largest group of 72 students scored in the range of 31-35. Only 14 students scored 20 or below.

Chart 8 shows the Difficulty Index according to the location of each item in the test. There is no clear pattern of item difficulty distribution throughout the test. The more difficult items are fairly evenly distributed throughout the test. The Pearson Correlation between Difficulty and Location (Item Number) is only 0.249.

There is no visually discernable relationship between Discrimination Indices and Location within the test as shown by Chart 9. The Pearson Correlation for this relation ship is only 0.028.

Chart 10 depicts the Discrimination Index compared to the Difficulty Index. A clear pattern is revealed. A Pearson Correlation between Difficulty and Discrimination Indices of –0.599 substantiates this visual observation. The higher (easier) the Difficulty Index, the less the item discriminates.

Chart 1

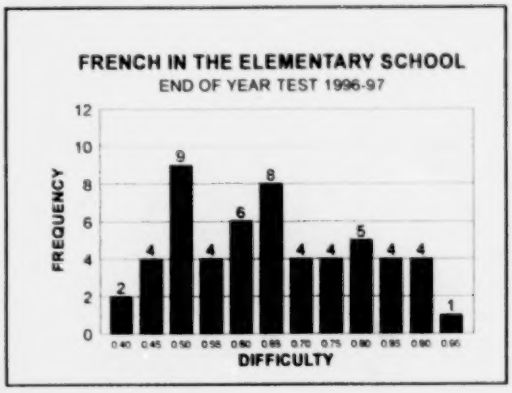


Chart 2

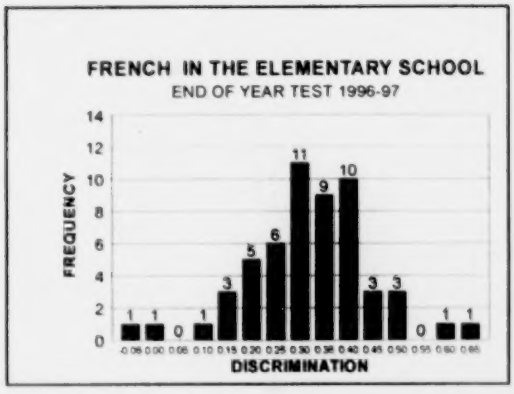


Chart 3

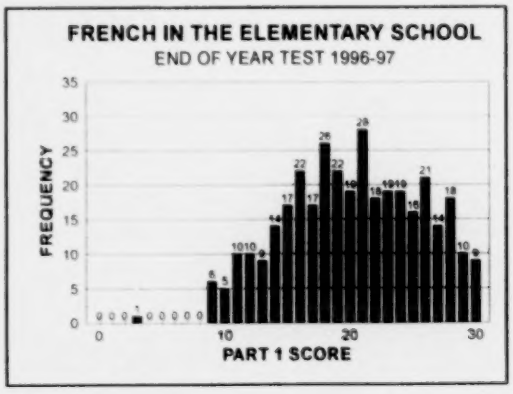


Chart 4

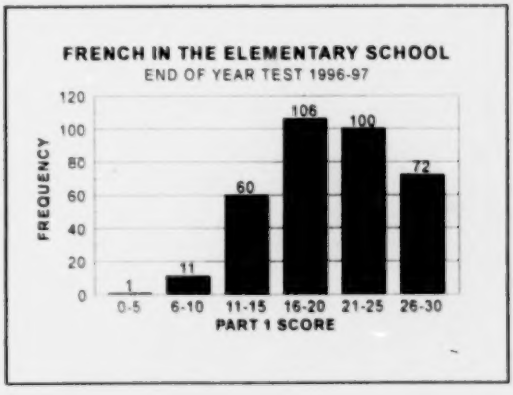


Chart 5

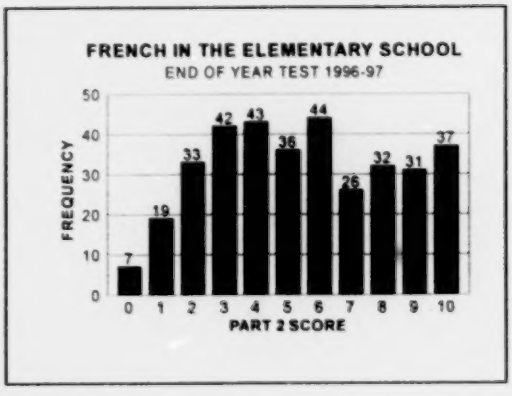


Chart 6

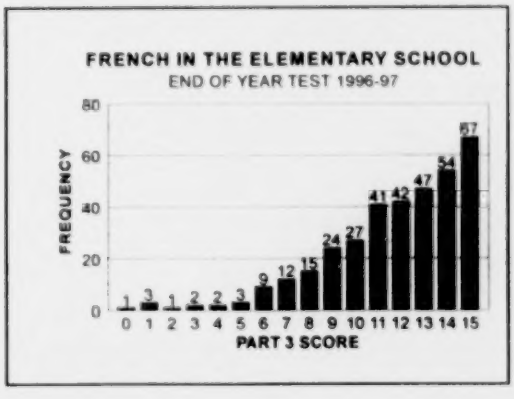


Chart 7

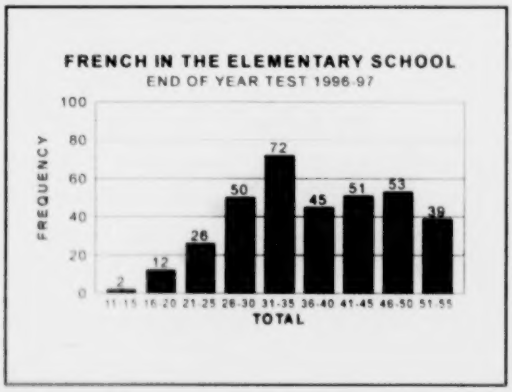


Chart 8

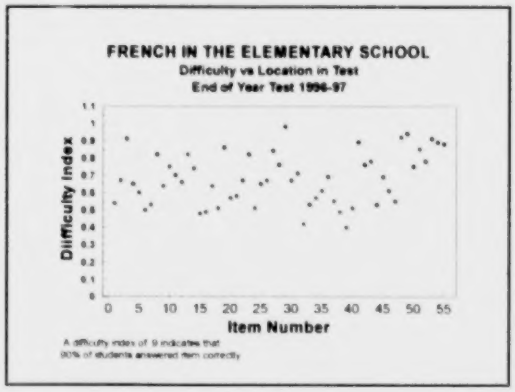

FRENCH IN THE ELEMENTARY SCHOOL
Difficulty vs Location in Test
End of Year Test 1996-97
Difficulty Index
0 5 10 15 20 25 30 35 40 45 50 55
Item Number
A difficulty index of 9 indicates that
90% of students answered item correctly

Chart 9

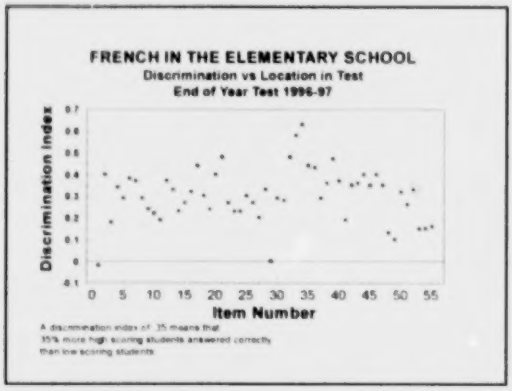
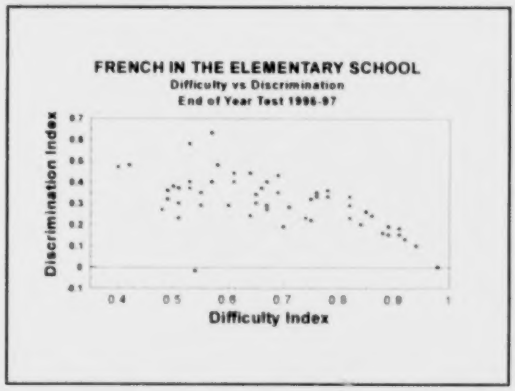

FRENCH IN THE ELEMENTARY SCHOOL
Discrimination vs Location in Test
End of Year Test 1996-97
Discrimination Index
0 5 10 15 20 25 30 35 40 45 50 55
Item Number
A discrimination index of .35 means that
35% more high scoring students answered correctly
than low scoring students

Chart 10


FRENCH IN THE ELEMENTARY SCHOOL
Difficulty vs Discrimination
End of Year Test 1996-97
Discrimination Index
0.4 0.5 0.6 0.7 0.8 0.9 1
Difficulty Index

Japanese

As in the case of the French Test, normalcy tests for Japanese indicated deviation from a normal distribution of scores at the .05 probability level. However, the deviations tended to be limited to the upper end of the distribution. The deviations were greater than those on the French Test.

The descriptive statistics for Japanese test results by subtest and total score are shown in Table 4. A total of 311 students took the Japanese test. The total scores ranged from 9 to 45 with three students scoring a zero on Part 2 and Part 3 of the test. The mean Total Score was 32.5.

Table 4: Summary Descriptive Statistics for Japanese by Subtest and Total

Variable	N	Mean	Median	Tr Mean	Min	Max	St Dev	SE Mean
Part 1	311	17.7	19	18.0	1	25	5.253	0.298
Part 2	311	6.9	7	7.0	0	10	2.321	0.132
Part 3	311	7.9	10	8.2	0	10	2.716	0.154
TOTAL	311	32.5	35	33.0	9	45	9.201	0.522

Correlations were obtained among Part 1, Part 2, Part 3 and Total Test scores. The parts are less related to each other than to the total test with a range of correlations between parts and the Total Test ranging from a low of 0.80 to a high of 0.95. Correlations among the parts ranged from 0.596 to 0.732. The high interrelationships between parts are to be expected because all parts of the Japanese Test involved listening skills. There is no reading part on the Japanese Test. These correlations are shown in Table 5.

Table 5: Correlations Among Parts and Total for Japanese

	Part 1	Part 2	Part 3
Part 2	0.652		
Part 3	0.732	0.596	
TOTAL	0.951	0.800	0.864

The Split-Half Correlation for the Total Test was 0.874 with a resulting Total Test Reliability of 0.932.

Difficulty and Discrimination Indices were computed for each item. These indices are found in Table 6. This information is also depicted graphically in Chart 11 and Chart 12. The Item Difficulty Index for Japanese ranges from 0.42 to 0.96. There are 33 test items, out of the 45, with a Difficulty Index of 0.70 or greater. Overall, there appear to be a large number of very easy items.

The Discrimination Index (Percentage of high-scoring group getting item correct – Percentage of low-scoring group getting item correct) ranges from 0.08 to 0.59. Nine test items have a Discrimination Index below 0.3 and one item has an index of 0.08.

Table 6: Japanese in the Elementary School
Item Difficulty and Discrimination

Item	Diff Index	Discrim Index	Item	Diff Index	Discrim Index	Item	Diff Index	Discrim Index
Part 1			16	0.96	0.08	6	0.68	0.21
1	0.42	0.32	17	0.78	0.33	7	0.64	0.46
2	0.56	0.40	18	0.88	0.23	8	0.67	0.35
3	0.85	0.32	19	0.80	0.33	9	0.74	0.30
4	0.79	0.39	20	0.73	0.45	10	0.77	0.40
5	0.50	0.44	21	0.70	0.34	**Part 3**		
6	0.74	0.43	22	0.69	0.52	1	0.80	0.39
7	0.82	0.34	23	0.76	0.30	2	0.71	0.52
8	0.76	0.35	24	0.65	0.25	3	0.76	0.45
9	0.62	0.38	25	0.75	0.29	4	0.66	0.59
10	0.58	0.32	**Part 2**			5	0.72	0.50
11	0.72	0.37	1	0.73	0.26	6	0.80	0.37
12	0.76	0.37	2	0.75	0.32	7	0.89	0.24
13	0.83	0.28	3	0.69	0.26	8	0.90	0.19
14	0.59	0.29	4	0.50	0.41	9	0.82	0.40
15	0.49	0.29	5	0.70	0.31	10	0.87	0.25

Score distributions for Part 1 are shown in Charts 13 and 14. Charts 15 and 16 depict distributions for Part 2 and Part 3. Chart 17 shows the Total Score Distribution by range of scores.

The scores for Part 1 ranged from 1 to 25 with 12 people making a perfect score. The distribution pattern for Part 1 is negatively skewed. When scores are grouped in 5-point interval ranges, 115 students scored between 21 and 25. Another 99 scored between 16 and 20.

On Part 2, 53 students made a score of 7 on the 10 questions. Of the 311 students taking the test, 190 scored in the 7-10 range. The distribution for Part 2 is also negatively skewed.

The Part 3 score distribution is extremely skewed with 158 students scoring 10 out of 10. Of 311 students tested, only 58 students scored 5 or below.

The Total test scores grouped by interval ranges of 5 tend to be negatively skewed. A total of 86 students scored between 36 and 40 on the 45-question test. Another 64 students scored 41-45. Only 6 students scored between 5 and 10, and 14 had a score between 11 and 15.

Chart 18 shows the Difficulty Index according to the location of each item in the test. It appears that the few relatively difficult items in the test were near the beginning of the test. Easy items appeared throughout the test. The Pearson Correlation between Difficulty and Location (Item Number) is 0.319.

The Discrimination Index by location in the test is graphically illustrated in Chart 19. Variability in Discrimination Indices is quite restricted up through Item 15. Items 16-22 show much wider variability. The scatterplot for Items 23-45 forms a "funnel" shape with a marked increase in variability across this range of item numbers.

Chart 20 depicts the Discrimination Index compared to the Difficulty Index. A clear pattern is again revealed. Although it appears less extreme from visual inspection than was the case with French, the Pearson Correlation is identically the same as it was for French: –0.345.

Chart 11

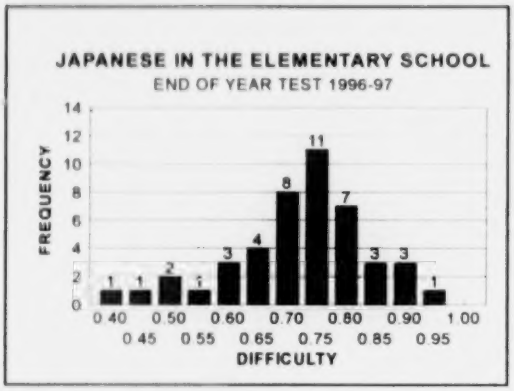

JAPANESE IN THE ELEMENTARY SCHOOL
END OF YEAR TEST 1996-97
FREQUENCY
DIFFICULTY
0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00

Chart 12

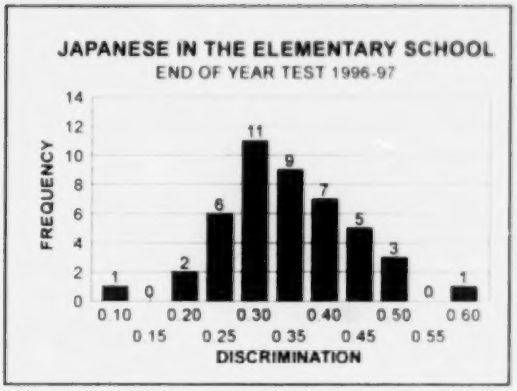

JAPANESE IN THE ELEMENTARY SCHOOL
END OF YEAR TEST 1996-97
FREQUENCY
DISCRIMINATION
0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60

Chart 13

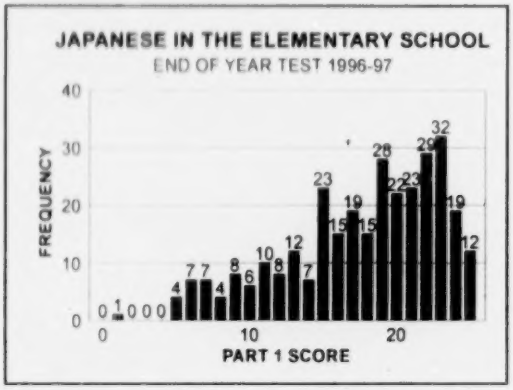


Chart 14

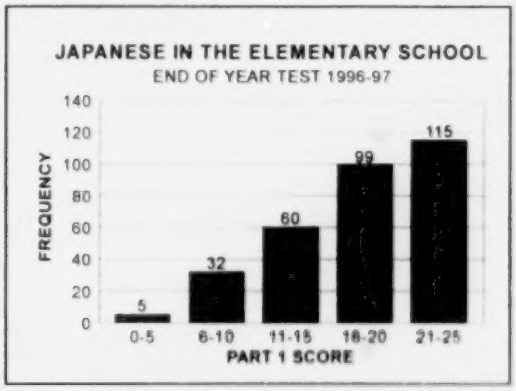


Chart 15

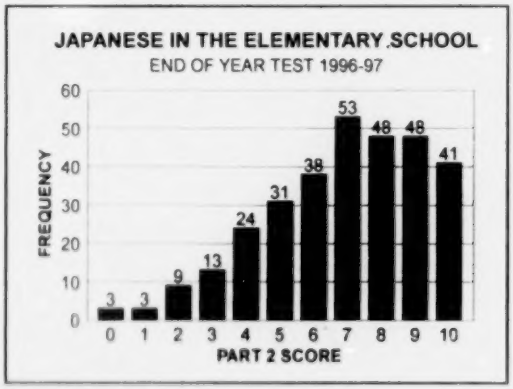


Chart 16

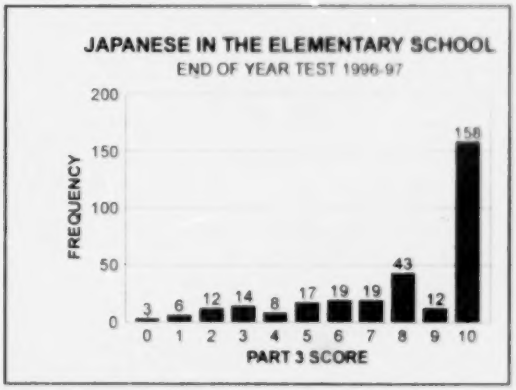


Chart 17

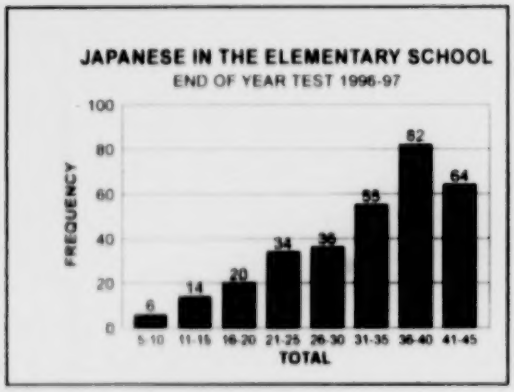


Chart 18

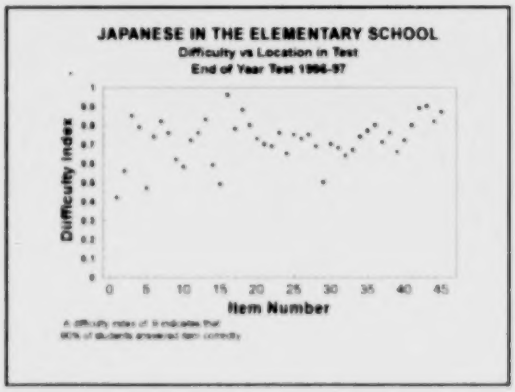


Chart 19

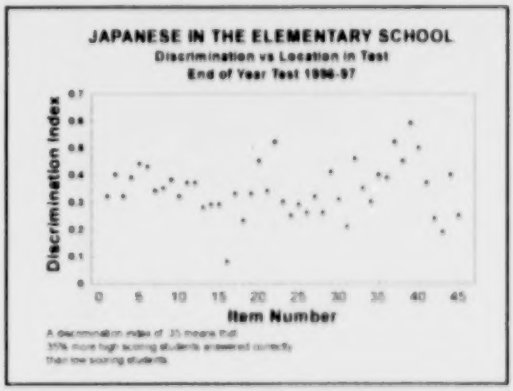
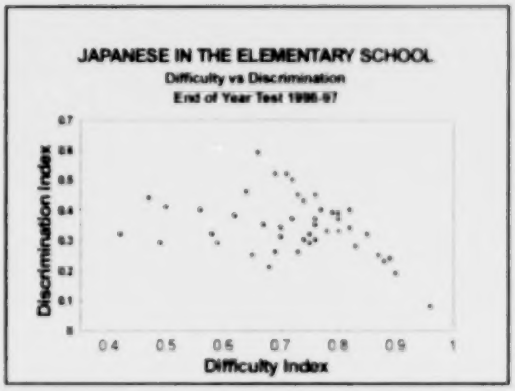


Chart 20



Spanish

Normalcy tests indicated deviation from a normal distribution of scores at the .05 probability level with the majority of the deviations occurring at the upper end of the distribution. The extent of the deviations was similar to the tests for the other two languages.

The descriptive statistics for Spanish test results by subtest and total score are shown in Table 7. A total of 434 students took the Spanish test. The total scores ranged from 14 to 55. One student scored a zero on Part 2 and Part 3 of the test. The mean Total Score was 38.9.

Table 7: Summary Descriptive Statistics for Spanish by Subtest and Total

Variable	N	Mean	Median	Tr Mean	Min	Max	St Dev	SE Mean
Part 1	434	22.1	23	22.4	7	30	5.284	.0.254
Part 2	434	6.1	6	6.1	0	10	2.570	0.123
Part 3	434	10.7	11	10.8	1	15	3.038	0.146
TOTAL	434	38.9	41	39.3	14	55	9.734	0.467

Correlations were obtained among Part 1, Part 2, Part 3 and Total Test scores. The inter-part and part-whole correlations tended to be slightly stronger for Spanish than for French but weaker than for Japanese with the correlations between the parts and the Total Test ranging from a low of 0.84 to a high of 0.94 and correlations among the parts ranging from 0.63 to 0.70. These correlations are summarized in Table 8.

Table 8: Correlations Among Parts and Total for Spanish

	Part 1	Part 2	Part 3
Part 2	0.702		
Part 3	0.688	0.630	
TOTAL	0.943	0.842	0.852

The Split-Half Correlation for the Total Test was 0.851 with a resulting Total Test Reliability of 0.919.

Difficulty and Discrimination Indices were computed for each item. These indices are found in Table 9. This information is also depicted graphically in Chart 21 and Chart 22. The Item Difficulty Index for Spanish ranges from 0.25 to 0.99. Thirty-four of the 55 items on the test have a difficulty index of 0.7 or higher. The test contains a large number of relatively easy items.

The Discrimination Index (Percentage of high-scoring group getting item correct – Percentage of low-scoring group getting item correct) range is 0.01 to 0.49. Nineteen test items have a Discrimination Index below 0.3. One item has a Discrimination Index of 0.01.

Table 9: Spanish in the Elementary School
Item Difficulty and Discrimination

Item	Diff Index	Discrim Index	Item	Diff Index	Discrim Index	Item	Diff Index	Discrim Index
1	0.64	0.13	20	0.83	0.25	8	0.58	0.41
2	0.57	0.34	21	0.61	0.39	9	0.71	0.39
3	0.95	0.12	22	0.63	0.45	10	0.73	0.48
4	0.88	0.21	23	0.47	0.40	**Part 3**		
5	0.77	0.31	24	0.57	0.31	1	0.87	0.21
6	0.79	0.16	25	0.74	0.44	2	0.80	0.34
7	0.59	0.34	26	0.49	0.34	3	0.74	0.40
8	0.90	0.21	27	0.93	0.13	4	0.93	0.14
9	0.54	0.29	28	0.76	0.38	5	0.79	0.25
10	0.78	0.34	29	0.99	0.01	6	0.80	0.28
11	0.76	0.30	30	0.74	0.43	7	0.57	0.37
12	0.83	0.25	**Part 2**			8	0.87	0.25
13	0.85	0.27	1	0.63	0.43	9	0.38	0.29
14	0.58	0.34	2	0.25	0.31	10	0.58	0.27
15	0.88	0.13	3	0.58	0.37	11	0.71	0.49
16	0.56	0.14	4	0.65	0.45	12	0.38	0.14
17	0.76	0.36	5	0.68	0.43	13	0.71	0.34
18	0.89	0.18	6	0.70	0.45	14	0.87	0.17
19	0.84	0.28	7	0.57	0.21	15	0.67	0.27

Score distributions for Part 1 are shown in Charts 23 and 24. Charts 25 and 26 depict distributions for Part 2 and Part 3. Chart 27 gives the Total Score Distribution by range of scores.

The range of scores for Part 1 is from 7 to 30 with 12 people scoring a 30. The distribution pattern is negatively skewed. When scores are banded into groups with 5-point ranges, 147 student scores are between 21 and 25, and another 135 scores are between 26 and 30, making a total of 282 students scoring from 21 to 30.

On Part 2, 214 of the 434 students scored between 7 and 10. The largest number of students, 68, obtained a score of 8.

The Part 3 score distribution is also negatively skewed. Of the 434 students taking the test, 263 scored in the range of 11 to 15 on this 15-question part. Only 15 students scored below 5.

The Total test scores exhibit a similar pattern. On the 55-question test, 221 students scored between 41 and 55. Only 21 students scored 20 or below.

Chart 28 depicts the Difficulty Index in relationship to location on the test. There is no clear pattern of item difficulty distribution among the 55 items based upon location within the test. The Pearson Correlation between Difficulty and Location (Item Number) is -0.155.

Like French and unlike Japanese, Chart 29 shows no obvious relationship between Discrimination Indices and location within the test. The Pearson Correlation between Discrimination and Location is 0.187.

Chart 30 compares the Discrimination Index and the Difficulty Index for each item graphically. The relationship shown is similar to that from the French and Japanese tests with a Pearson Correlation between Difficulty and Discrimination Indices of –0.377.

Chart 21

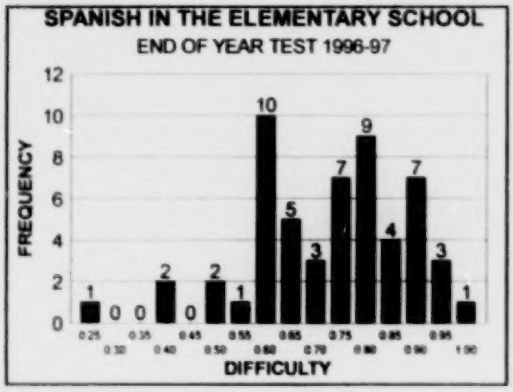


Chart 22

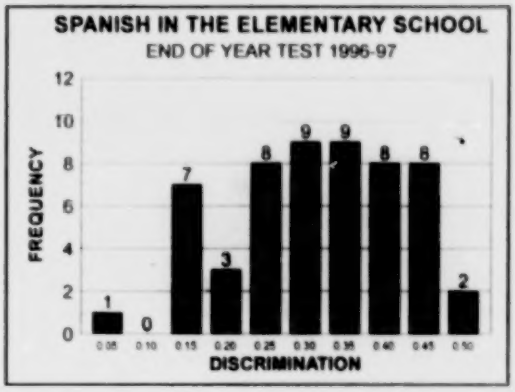


Chart 23

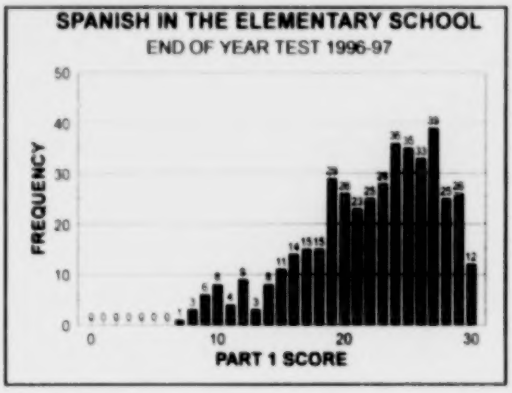


Chart 24

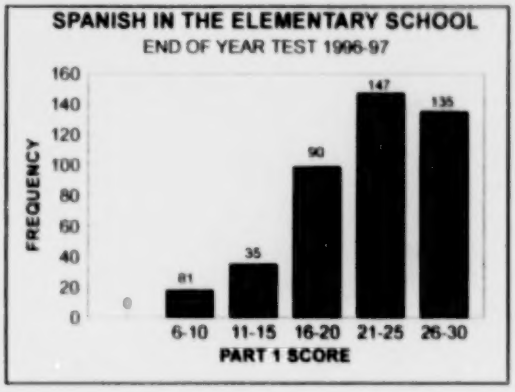


Chart 25

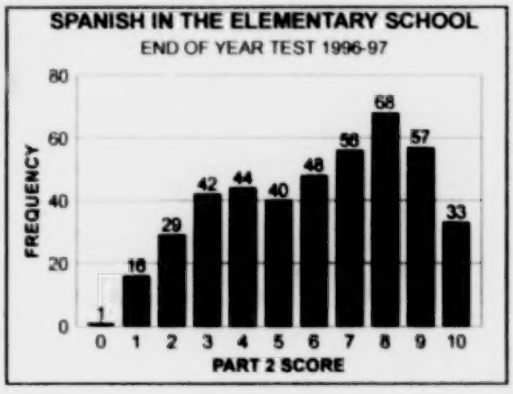


Chart 26

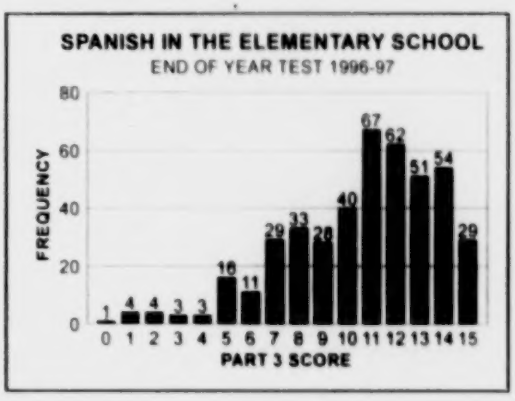


Chart 27

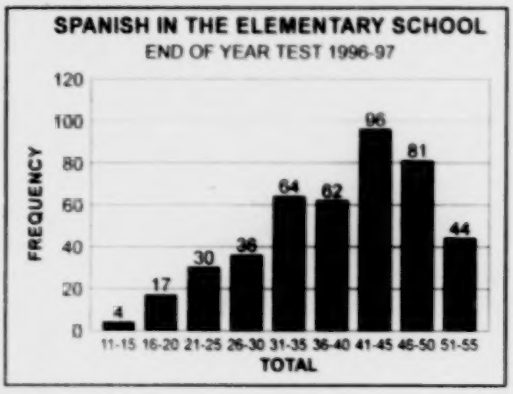


Chart 28

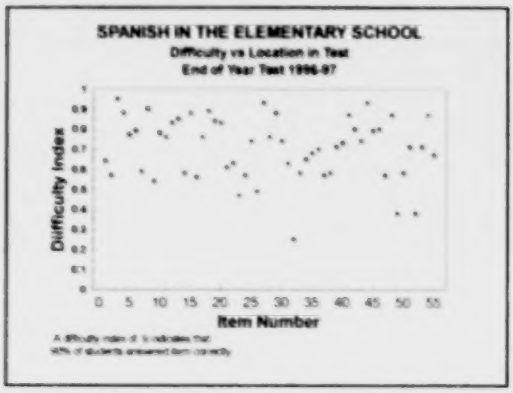

SPANISH IN THE ELEMENTARY SCHOOL
Difficulty vs Location in Test
End of Year Test 1996-97
Difficulty Index
0 5 10 15 20 25 30 35 40 45 50 55
Item Number

Chart 29

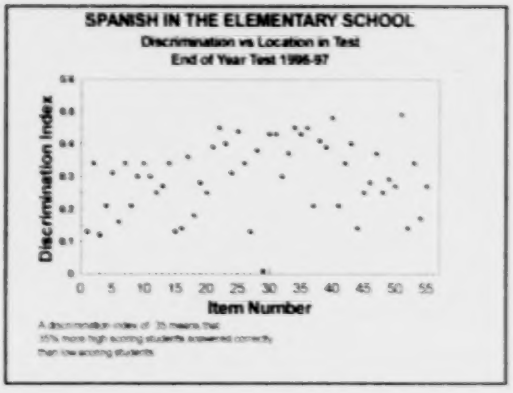
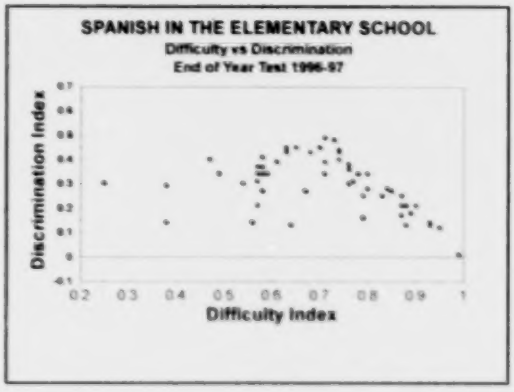

SPANISH IN THE ELEMENTARY SCHOOL
Discrimination vs Location in Test
End of Year Test 1996-97
Discrimination Index
0 5 10 15 20 25 30 35 40 45 50 55
Item Number

Chart 30


SPANISH IN THE ELEMENTARY SCHOOL
Difficulty vs Discrimination
End of Year Test 1996-97
Discrimination Index
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Difficulty Index

Chapter 5

Needed Revisions To Testing Program

The district's overall goal for its foreign language assessment instruments is to produce test results that give accurate pictures of achievement levels of entire groups of students tested. Test items selected for each test should be of appropriate difficulty and discrimination levels so that the resulting test scores approximate a normal distribution. There should be sufficient numbers of both easy and difficult items to insure that both low and high achieving students have the opportunity to demonstrate their true achievement levels. Few, if any, students should miss all of the items or get all of the items correct.

With this goal in mind, the following revisions are recommended:

> Review all test items that 75% or more of the students answered correctly. There are 14 French, 14 Japanese and 24 Spanish items in this category. Most of these items need to be revised or replaced. The negatively skewed distributions, indicating a preponderance of easy items in these tests, suggest the need for elevating teacher expectations for student achievement in foreign language at the elementary level.

> Review all test items that have a discrimination index of 0.10 or less. There are 3 French, 1 Japanese and 1 Spanish item in this category. These items should be revised or replaced. There are an additional 7 Spanish items with a discrimination index of less than

0.15 that should receive close scrutiny during the next testing cycle.

Organize formal ongoing processes for the development of parallel test items to be included in item banks, field testing of items and selection of items to be included in the annual test administration for each foreign language in accordance with Tables of Specifications.

Timely implementation of these recommended revisions is extremely important. Specifically, item revisions need to be completed and revised items included in the Spring 1998 end-of-year testing.

Recommendation for Further Study

The dearth of quality assessment instruments for foreign language in elementary school programs places the evaluation program being developed by this school district in a favorable position for adoption by a wide range of school districts. All analyses completed to date have been based upon Classical Testing Theory. The use of Classical Testing Theory (Crocker & Algina, 1986) poses no problem for utilization within the district which has a highly homogeneous student population. However, since item difficulty and discrimination are a function of the sample utilized, these item characteristics may not be generalizable to other populations which may differ substantially from the student population involved in this study. It is, therefore, strongly recommended that Item Response Theory (Hambleton, Swaminathan, & Rogers,

1991) be used to analyze the test items. This analysis will provide item characteristics independent of the sample population and will increase potential applications many fold.

Selection of an appropriate IRT model is extremely important. If only item difficulty were to be considered, a Rasch One-Parameter Model (Andrich, 1988) would be appropriate. However, item difficulty and item discrimination are both pertinent considerations. Since the instruments utilize multiple-choice items, guessing becomes a factor which probably should be represented in the model in order to obtain a good fit between the model and the data.

Several computer programs are available today for parameter estimation in IRT Models. LOGIST (Wingersky, Barton, & Lord, 1982) fits one-, two-, and three-parameter models using joint maximum likelihood estimation. BILOG (Mislevy & Bock, 1984) also uses joint maximum likelihood procedures but allows for optional Bayesian procedures. Software selection has been further complicated by a recent proliferation of less well-known computer programs for IRT models. Such selection should only be made in consultation with an experienced user of a variety of IRT parameter estimation software packages.

References

Andrich, David. (1988). Rasch Models for Measurement. Newbury Park, CA: Sage Publications.

American Council on the Teaching of Foreign Languages. (1986). ACTFL Proficiency Guidelines. Hastings-on-Hudson, New York: American Council on the Teaching of Foreign Languages.

Bachman, L. F. (1990). Fundamental Considerations in Language Testing. Oxford: Oxford University Press.

Crocker, Linda and Algina, James (1986). Introduction to Classical & Modern Test Theory. San Antonio: Harcourt Brace Jovanovich College Publishers.

Donato, Richard. (1996). Monitoring and Assessing a Japanese FLES Program: Ambience and Achievement. Washington, DC: Center for Applied Linguistics. (ERIC Document Reproduction Service No. ED 397 632).

Foreign Language and Youth. (1996, December 18). Mountain Democrat, Editorial Page (http://mtdemocrat/editorials/foreign).

Gollnick, D. and Chinn, P. (1994). Multicultural Education in a Pluralistic Society. Riverside, New Jersey: Macmillan.

Hambleton, Ronald K., Swaminathan, H. and Rogers, H. Jane. (1991). Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications.

Hancock, Mary. (1995). Student Perceptions of an Elementary Foreign Language Curriculum. Washington, DC: Center for Applied Linguistics. (ERIC Document Reproduction Service No. ED 392 249).

Hanley, Julia E. B. (1995, Spring). Using Video as an Advanced Organizer to a Written Passage in the FLES Classroom. Modern Language Journal, 79(1), 57-66.

Koren, Shira. (1995, August). Foreign Language Pronunciation Testing: A New Approach. System, 23(3), 387-400.

Learning at an Early Age. (1997, September 8). Times Record, Editorial Page (http://www.timesrecord.com/editorial).

Mislevy, R. J. and Brock, R. D. (1984). BILOG: Maximum Likelihood Item Analysis and Test Scoring with Logistic Models. Mooresville, IN: Scientific Software.

North Carolina Department of Public Instruction. (1995, April). Chapel Hill/Carrboro Second Language Program. Raleigh, NC: Information Webb, North Carolina Department of Public Instruction (http://oaktree.dpi.state.nc.us).

The National Standards in Foreign Language Education Project. (1996). Standards for Foreign Language Learning: Preparing for the 21st Century. Lawrence, Kansas: Allen Press.

Vivas, Eleonora. (1996, June). Effect of Story Reading on Language. Language Learning, 46, 189-216.

Wingersky, M. S., Barton, M. A. and Lord, F.M. (1982). LOGIST User's Guide. Princeton, NJ: Educational Testing Service.